INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad - 500 043

MODEL QUESTION PAPER-I

B.Tech VII Semester End Examinations, DEC–2023

Regulations: IARE - UG20

PREDICTIVE DATA ANALYTICS
COMPUTER SCIENCE AND ENGINEERING(AI&ML)

Time: 3 hour   Maximum Marks: 70

Answer ONE Question from each MODULE
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

MODULE-I

1. (a) Do the tools provide ways to interpret and explain the predictions made by the models? How can we ensure that the models' predictions are transparent and understandable to stakeholders? [**BL:Analyze—CO1—Marks:7**]

   (b) Explain the importance of data visualization in conveying insights to stakeholders.What principles should be followed to create effective data visualizations for decision-makers?Can you share examples of impactful data visualizations that influenced critical decisions. [**BL:Analyze— CO1—Marks:7**]

MODULE-II

2. (a) In what situations is it crucial for models to be interpretable and explainable, and how can organizations achieve this? What techniques can be employed to make complex machine learning models more understandable to non-technical stakeholders?

   [**BL:Analyze— CO2—Marks:7m**]

   (b) Discuss the ethical challenges that may arise during the data-to-insights-to-decisions process, such as bias, privacy, and transparency. Explain how organizations can address these challenges responsibly.

   [**BL:Understand— CO2—Marks:7m**]

MODULE-III

3. (a) When selecting a machine learning model for information-based learning, what factors should be considered, and how can one determine the most suitable model for a given problem?

   [**BL:Understand— CO3—Marks:7m**]

   (b) Explain the basic concept of the ID3 algorithm for decision tree construction. Discuss its objectives and significance in machine learning and data analysis. Provide a high-level overview of the key steps involved in ID3.

   [**BL:Analyze— CO3—Marks:7m**]

4. (a) In multiclass classification, how is Shannon entropy extended to measure the impurity or disorder in a set of classes? Discuss its role in algorithms like decision trees.

   [**BL:Analyze— CO3—Marks:7m**]

   (b) Discuss different similarity metrics commonly used in similarity-based learning, such as Euclidean distance, cosine similarity, Jaccard similarity, and Pearson correlation coefficient. Explain the mathematical formulations of these metrics and provide scenarios where each is appropriate.

   [**BL:Understand— CO3—Marks:7m**]

MODULE-IV

5. (a) Explain the fundamental concept of Bayes' Theorem. Provide the mathematical formula and discuss the terms involved, such as prior probability, likelihood, marginal likelihood, and posterior probability. How does Bayes' Theorem help in updating beliefs or probabilities?

[BL:Understand— CO4—Marks:7m]

(b) In a decision-making context, how can Bayes' Theorem help you make optimal decisions by considering probabilities, costs, and benefits? Provide an example where Bayesian decision-making is applicable.

[BL:Understand— CO4—Marks:7m]

6. (a) Explain the concept of conditional independence in probability theory. How does it differ from unconditional independence? Provide an example that illustrates the difference between these two types of independence.

[BL:Understand— CO4—Marks:7m]

(b) You are developing a spam email filter. Explain how Bayesian prediction, specifically using the Naive Bayes classifier, can be used to classify incoming emails as spam or not. What are the key features and probabilities involved in this prediction?

[BL:Apply— CO4—Marks:7m]

MODULE-V

7. (a) In a supervised learning scenario, you have a classification model that frequently misclassifies a specific class. How would you analyze the misclassified instances to understand the sources of error?
[BL:Analyze— CO5—Marks:7m]

(b) Explain the Concept of Error Measurement in Statistics and Its Significance in Data Analysis.

[BL:Evaluate— CO5—Marks:7m]

8. (a) Explain the Concept of Bias and Variance in Error Measurement and Their Trade-off in Model Selection?

[BL:Understand— CO5—Marks:7m]

(b) Select a specific industry or domain (e.g., healthcare, marketing, finance) and discuss how multivariable linear regression with gradient descent is commonly used to address practical problems in that domain.
[7m]

[BL:Understand— CO5—Marks:7m]

END OF EXAMINATION

COURSE OBJECTIVES:
The course should enable the students to:

1	The fundamental concepts of data analytics.
2	The principles and methods of statistical analysis.
3	The models used in predictive data analytics applications using supervised machine learning.
4	The interesting patterns analyze supervised and unsupervised models and estimate the accuracy of the algorithms.

COURSE OUTCOMES:
**After successful completion of the course, students should be able to:**

CO 1	**Demonstrate** the fundamental concepts and types of machine learning its role in predictive data analytics and real-world applications	Understand
CO 2	**Make use of** descriptive and exploratory analysis methods to extract valuable insights from data to provide a foundation for decision-making.	Apply
CO 3	**Build** a techniques to collect relevant data and assess its quality ensuring the foundation for accurate information extraction	Apply
CO 4	**llustrate** probability-based learning to develop probabilistic classification and regression models) capturing uncertainty in predictions.	Understand
CO 5	**Utilize** the principles of error-based learning and its significance in improving models through the analysis of prediction errors.	Apply
CO 6	**Identify** the foundational concepts of machine learning, including types of algorithms, training, and prediction (and their application in predictive data analytics	Apply

QUESTION PAPER 1: MAPPING OF SEMESTER END EXAMINATION QUESTIONS TO COURSE OUTCOMES

Q.No		All Questions carry equal marks	Taxonomy	CO's	PO's
1	a	Do the tools provide ways to interpret and explain the predictions made by the models? How can we ensure that the models' predictions are transparent and understandable to stakeholders?	Analyze	CO 1	PO 1

	b	Explain the importance of data visualization in conveying insights to stakeholders.What principles should be followed to create effective data visualizations for decision-makers?Can you share examples of impactful data visualizations that influenced critical decisions .	Analyze	CO 1	PO 1
2	a	In what situations is it crucial for models to be interpretable and explainable, and how can organizations achieve this? What techniques can be employed to make complex machine learning models more understandable to non-technical stakeholders?	Analyze	CO 2	PO 1
	b	Discuss the ethical challenges that may arise during the data-to-insights-to-decisions process, such as bias, privacy, and transparency. Explain how organizations can address these challenges responsibly.	Understand	CO 2	PO 1
3	a	When selecting a machine learning model for information-based learning, what factors should be considered, and how can one determine the most suitable model for a given problem?	Understand	CO 3	PO 1
	b	Explain the basic concept of the ID3 algorithm for decision tree construction. Discuss its objectives and significance in machine learning and data analysis. Provide a high-level overview of the key steps involved in ID3.	Analyze	CO 3	PO 1, PO 2
4	a	In multiclass classification, how is Shannon entropy extended to measure the impurity or disorder in a set of classes? Discuss its role in algorithms like decision trees..	Analyze	CO 3	PO 1
	b	Discuss different similarity metrics commonly used in similarity-based learning, such as Euclidean distance, cosine similarity, Jaccard similarity, and Pearson correlation coefficient. Explain the mathematical formulations of these metrics and provide scenarios where each is appropriate.	Understand	CO 3	PO 1, PO 2

5	a	Explain the fundamental concept of Bayes' Theorem. Provide the mathematical formula and discuss the terms involved, such as prior probability, likelihood, marginal likelihood, and posterior probability. How does Bayes' Theorem help in updating beliefs or probabilities?	Understand	CO 4	PO 1, PO 2
	b	In a decision-making context, how can Bayes' Theorem help you make optimal decisions by considering probabilities, costs, and benefits? Provide an example where Bayesian decision-making is applicable.	Understand	CO 4	PO 1, PO 2
6	a	Explain the concept of conditional independence in probability theory. How does it differ from unconditional independence? Provide an example that illustrates the difference between these two types of independence.	Understand	CO 4	PO 1, PO2
	b	You are developing a spam email filter. Explain how Bayesian prediction, specifically using the Naive Bayes classifier, can be used to classify incoming emails as spam or not. What are the key features and probabilities involved in this prediction?	Apply	CO 4	PO 1, PO2
7	a	In a supervised learning scenario, you have a classification model that frequently misclassifies a specific class. How would you analyze the misclassified instances to understand the sources of error.	Analyze	CO 5	PO 1, PO 2
	b	Explain the Concept of Error Measurement in Statistics and Its Significance in Data Analysis.	Evaluate	CO 5	PO 1, PO 2, PO 3
8	a	Explain the Concept of Bias and Variance in Error Measurement and Their Trade-off in Model Selection?	Understand	CO 6	PO 1, PO 2, PO 12
	b	Select a specific industry or domain (e.g., healthcare, marketing, finance) and discuss how multivariable linear regression with gradient descent is commonly used to address practical problems in that domain.	Understand	CO 6	PO 1, PO 2, PO 3

KNOWLEDGE COMPETENCY LEVELS OF MODEL QUESTION PAPER

BLOOMS TAXONOMY
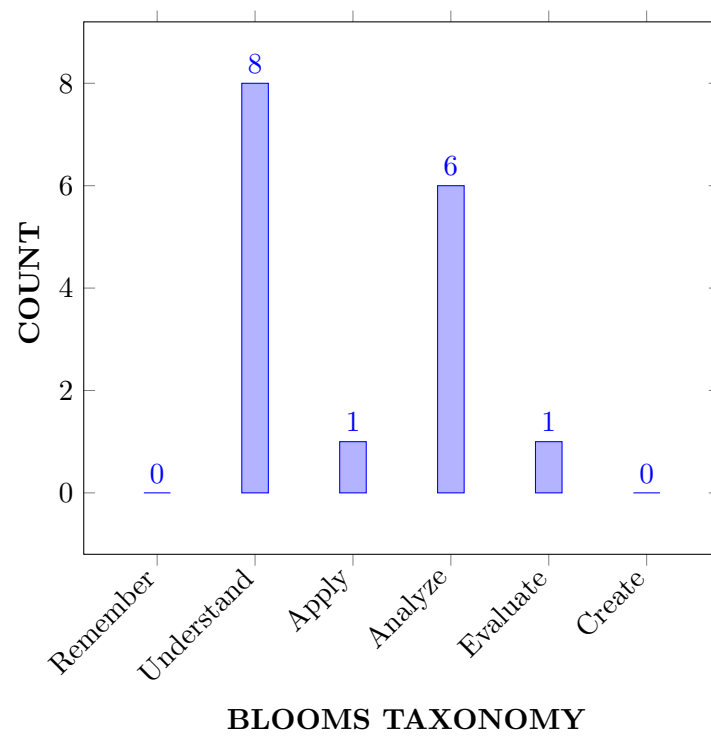
Signature of Course Coordinator   **HOD,CSE(AI&ML)**